

Concurrent Session One -- Data Management

David Cotton
Facilitator
ORC Macro

David Cotton welcomed the group to the session, noting that it was an opportunity to learn from each others' challenges and successes. He encouraged everyone to share issues and feedback, as well as ways to address those issues. He said that suggestions for features of data management systems would be helpful. These features might include the kinds of reports generated and the nature of interfaces. He posed the question, "If you had unlimited resources, what kind of system would you want?"

Choi Wan
CDC Representative
CDC/PERB

Choi Wan updated the group on the ERAS system, indicating that it is an aggregate reporting system. Eight to nine different health departments will be pilot-testing the system. They hope that the second phase will be even more helpful. He then described another upcoming product – health department software. Local health departments can use this software to organize their CBO or funding agency data to use the data as well as aggregate it for CDC. The system has been delayed, but the time-frame for the software is a six-to-nine month turnaround. By then, the system will be available to health departments, as will training and technical support.

Mari Gasiorowicz
Health Department Peer
State of Wisconsin Health Department

Mari Gasiorowicz said that in Wisconsin, their approach to data collection and management is different from other states. They have standardized forms for everything from intervention plans to data collection forms, and they require all of their agencies to use these forms, placing a heavy burden on them to prepare the intervention plans, decide which type of intervention plan to use, to collect the data, and to enter the data. They are in the pilot-testing phase of a web-based system. Their philosophy and goal is for agencies to be able to use and manipulate their data, increasing buy-in and ownership of the intervention plan and the data, at every level from administrative to outreach. They are also committed to a significant amount of training and technical assistance.

They have several parts to their data collection and management system. They incorporated the

CDC Evaluation Guidance into their HIV prevention training and interventions. Within the intervention plan manual, there are seven intervention plans, and training is available either to providers or to the population. She offered examples of their data collection and reporting forms, which their agencies complete as part of their intervention plans. There are tables to describe target populations, including the total number of clients served, race and ethnicity, and HIV-status, if available. Agencies make their expected target populations part of their intervention plan – the basics of who they think that they will reach pertain to data management, especially in anticipated outcomes. Agencies project how many people they think they will see.

The intervention plan is an important piece of the work. She produced a red binder, which contained data collection forms, samples, and instructions. Data is collected on the group and individual levels. In Wisconsin, they collect client-level data. Many states aggregate data, while others keep it at the individual level. The forms in the red binder correspond to the website. Instead of a client code, they use a provider-based code, which includes the initial of a staff person. This code follows the client within an agency. Client information includes the initial date of service, the source of services, counseling and testing, and other data.

The website can also track prevention case management (PCM) data, noting how many clients participate in a given intervention plan: each intervention plan has a different code. The use of a provider-based code for each client eliminates the need to complete client information more than once: details such as gender, age, race, and ethnicity are held in the client code. There is also space in the forms to note referrals to other services, when the referral(s) took place, whether information was provided or the referral visit was more directly facilitated, and whether it was completed.

The intervention plan tracking also includes services tracking, so it is noted if clients belong to more than one intervention plan. Due to their funding in the state, she noted that they keep more detailed information in this area than other states might have to keep. Each session is committed to the system, including the date of contact, the amount of time spent, the kind of time spent (for instance, face-to-face), the setting, whether incentives were provided, whether goals were set or reviewed. The modules and topics covered are also noted.

In the web-based system, agencies enter their data by intervention plan code, both at the client level and for intervention services tracking for ILI, GLI, and PCM. She then demonstrated the Internet site, asking the group for feedback, as it was still in the pilot-testing phase.

Agencies can enter new intervention plan types, and the forms correspond to the paper forms given by the health department. Intervention plan data includes funding details, such as funding sources and total clients served. Mari Gasiorowicz told the group that there was no way, at present, to sort the clients in any way, such as by client ID or point of entry. As the system is web-based, there is no limit to how many users can use the system at the same time.

She spoke about the department's decision to code their clients using a provider-based code. This system relates more to provider perception than to actual confidentiality issues. Provider codes reassure participants that they are anonymous, despite the many good codes that incorporate a piece of the client's name or use another method for tracking clients across agencies.

Most interesting to their agencies is the report for ILI, GLI, and PCM. This report compares predictions to actual results of the work and tracks programs' progress toward their goals. With this feature, grantees know their status exactly, including client demographics and referrals. The kinds of interventions being conducted are also tracked. Adding narrative is possible as well. Agencies are responsible for logging all of their data by thirty days after the end of the quarter, including a narrative for each intervention plan.

Their grantees have not begun to enter real data, as they are still in the pilot-testing phase. Training and technical assistance on both the intervention plan and the data collection forms have been an important aspect of the health department's work to get the system on-line. They have improved intervention plans, they expect to get better data, and their agencies seem to be committed to the system.

In conclusion, Mari Gasiorowicz gave the group a list of some of her lessons learned from the process:

- ☐ The project has taken longer than they thought it would, and they are not done yet;
- ☐ The testing phase is critical; and
- ☐ Having the paper forms ahead of time was helpful so that agencies could become familiar with them and use them to collect data for subsequent entry into the web-based system.

Discussion Summary:

- ❖ An audience member asked Mari Gasiorowicz about the cost of the system.
- ❖ Mari Gasiorowicz replied that the state had a bioterrorism grant under which they are developing a health alert network system. That project has incorporated their reporting system. She asked their web developer to estimate his time spent on the site, and he guessed that his hours of work would total approximately \$10,000. Development costs were minimal for them, but the project has taken a great deal of her time and the state epidemiologist's time.
- ❖ Another audience member asked about the intervention forms.

- ❖ Mari Gasiorowicz said that when people are doing their interventions, they use the state's standardized forms. There are different forms for clients, services, and outreach.
- ❖ Another participant inquired as to how their web system related to CDC's ERAS system.
- ❖ Mari Gasiorowicz said they have an ACCESS database, but she also understood that web-based systems are not optimal for detailed cross-tabs. They manipulate the web-gathered information separately, which allows them to generate CDC reports, which then go into the ERAS system.
- ❖ An inquiry was posed about the number of providers and grantees that Wisconsin has and about how they feel about the system.
- ❖ Mari Gasiorowicz replied that they have about thirty grantees. The time involved has been a problem to some of the pilot-testers, but they are learning how to enter the data fast.
- ❖ Another audience member commented that New Jersey has been collecting client-level data for a while. States are in multiple stages, but he urged them not to collect data just for CDC purposes. Their forms include each agency's needs, the state needs, and the CDC needs. With this approach, the agencies get information that they can use and that also satisfies CDC's reporting requirements. Before choosing a system, he stressed that it is important to work with CDC to ensure data compatibility.
- ❖ A participant from a smaller state commented that they do not have the in-house personnel resources to draw on to implement these systems. In small states, the three or four people in their office will be putting these systems on-line. He expressed his hope that CDC would remember that states have different levels of resources.
- ❖ Another participant added that ongoing technical support for the CBOs is a real issue. Where they expected a need for assistance in using the software, they have found that the needs for assistance are coming in the program evaluation, including definitions and guidance.
- ❖ Hope Cassidy-Stewart commented that in Maryland, they were not web-based. The core of any system seems to be helping CBO's understand it and how it relates to them and their projects. The quality of the data depends on the people in the field, and training them takes a lot of work.
- ❖ David Cotton added that there are several layers to the work, from working with contractors to assuring quality data input to the issue of the data management system

itself.

- ❖ A participant said that his state gives reports twice a year and makes comments on progress. Sharing this information with grantees interactively would be a great asset, creating an ongoing dialogue between the progress monitors and the grantees.
- ❖ A participant from Massachusetts noted that there are eleven data collection systems being used. She wondered about a possible forum for sharing IT-level information.
- ❖ A participant from Minnesota said they found that getting information from the non-technical, evaluation staff to the IT staff was impossible. He recommended that CDC do periodic video conferences for technical people so that they can keep up-to-date on changes.
- ❖ Hope Cassidy-Stewart suggested that they set up their own conference calls to share information.

Jaime Altamirano
CDC Representative
Technical Assistance Opportunities

Jaime Altamirano said that when a new system is initiated, even if it is designed to improve on an existing system, it is expected that the new system will conflict or clash with the current system. Therefore, as soon as the new Guidelines came, they were prepared for the number of technical assistance requests that came in. The requests came from four aspects of data management:

- ☐ Generating data: reporting aggregated data to CDC may require changing the states' system of data collection, creating new forms, changing the data collection methods on the state level;
- ☐ Data entry;
- ☐ Data validation; and
- ☐ Transferring data back to CDC.

He reinforced CDC's understanding of the challenges that the changes brought about. Taxonomy and interpretation of the CDC guidelines were the bulk of initial TA requests. Then, requests centered around reconciling the new forms with extant state forms and requirements. Some states had completed development of their own data collection system and were being

asked to modify it again. This frustration could sometimes be alleviated by understanding the benefit of a universal method of data collection, which facilitates comparisons of data at the state and national levels.

Instead of dwelling on definitions, Jaime Altamirano focused on the action steps required to report data according to the new guidelines.

- ☐ Adaptation to the new forms or the creation of new forms at the state level
- ☐ Changes in data collection methods
- ☐ Data entry and validation of data
- ☐ Data analysis and reporting of aggregated data

State health departments are not expected to conduct in-depth analysis on data, but only to report aggregate data to CDC. However, he urged them to consider how their data can be used for their benefit and for the benefit of their providers and CBOs. They should anticipate further data analysis as they develop their databases.

Data transfer is the last step, and CDC is creating the ERAS system for this reporting. Some states are not concerned about the data transfer process, as the first steps are more important to them. He pointed to Wisconsin's process as a good first step in developing a system for data collection and reporting. Other states are still waiting for CDC's assistance or the software. When a request for technical assistance comes, CDC adapts to the needs and technology levels of each state. Requests come through Project Officers to the Science Application Team.

CDC's software is anticipated in six to nine months. In the meantime, the states and jurisdictions who have developed, or are in the process of developing, their own models, are willing to share ideas about their systems. They can also share experiences, barriers, or limitations in the process. CDC will create an avenue to share that information from state-to-state. Hardware requirements are a capacity-building request, he noted, but he expected that most concerns were with software and with training. Health departments have to collect data from CBOs, taking into account the upcoming CBO guidelines. The system must be compatible with these guidelines and ensure that the CBO and local-level data will be reported to the state level. Each state has its own concerns in this area, so the technical assistance team analyzes each situation when providing technical assistance.

Discussion Summary:

- ❖ A participant asked, if the software that CDC is developing is similar to Wisconsin's system but not the same, whether they would accomplish similar goals.
- ❖ Jaime Altamirano indicated that they would accomplish similar goals. They are all trying

to develop a system that will help the state collect the right data into the right database that will generate useful aggregate data.

- ❖ Another participant inquired whether the ERAS system would have a place for a prevention plan.
- ❖ David Cotton replied that the system does have a place for intervention plan data, but it is not connected to goals or process monitoring data. Data linkage is included at the health department aggregate level, not at each CBO or each intervention level. The new system will address the relationship between the provider and the health department where ERAS is concerned with the transfer of information between the health department and CDC.
- ❖ An inquiry was posed as to whether the ERAS system had space for client-level data, and a way to enter information about an intervention plan for each of the state's agencies. Will the system aggregate the data and send it to CDC?
- ❖ David Cotton pointed out that the new software in development will provide aggregates at the health department level from either client-level or intervention-level data.
- ❖ Since development of systems takes longer than expected, an inquiry was posed as to CDC's best case scenario of when the software would be available, as opposed to the estimate of six to nine months.
- ❖ Jaime Altamirano replied that their difficulty was in creating one product that will apply to several states, given the different needs of each individual state. David Cotton agreed that six to nine months is, at the least, optimistic.
- ❖ A participant asked about the ability of the system to store electronic data since databases can take a lot of space, and systems can crash if they are overloaded.
- ❖ Jaime Altamirano said he appreciated the importance of the issue. He noted that access is also a concern. Having data entry at different points makes validation very difficult because of different capacities at different sites.
- ❖ Tim Juday, from Hawaii, asked what health departments should do in the time before the ERAS system is up and available. He wondered whether they should work to develop their own web-based system, or whether they should invest their own resources in developing their own systems. He also commented that the data collection system and the ERAS are illogical from a statistician's point of view. Each locality collects its own data, and when national-level is aggregated, problems are likely to occur with each

state's own way of collecting data and its own definitions, regardless of what the CDC says. There are going to be validity and viability problems, and statistical analyses with those data will not mean much, so that policy that comes out of those analysis will be questionable.

- ❖ Choi Wan replied that CDC would support states that wanted to develop their own data collection systems. They are, however, asking states that might not have that capacity to hold off on creating their own systems until they can be more certain of when the ERAS system will come on-line. When PGO makes an official, clear announcement, then the turnaround time will be very quick, he said. In short, if a state has begun working on a system, then that state should continue. If a state has not begun, then they might wait. He said that the six-to-nine month time-frame was his hope, and that it represented not only when the system would be ready, but also when technical assistance would be available. Choi Wan acknowledged that there were fully aware of the issue of different health departments reporting different data and using different definitions. This is why no aggregate data has been released to date. They do, however, want to test some mechanisms to unify definitions and systems, hoping that in the future, health departments will come to agreement about definitions. Aggregating information across states and departments makes sense not only for CDC, but also for local governments. At this moment, Choi Wan agreed that there is a danger in aggregating data.
- ❖ Jaime Altamirano added that states should pursue their own interests in data collection so that they might manage their data for themselves and their localities. It will be beneficial to be compliant with the CDC software as well, so he encouraged states to keep both their local needs and CDC requirements in mind.
- ❖ An inquiry was posed as to when the RFP for data systems would be released, and for what period of time that funding would be available.
- ❖ Choi Wan answered that they hoped that the announcement would come very soon, maybe by July, 2001.
- ❖ An inquiry was posed as to whether CDC is using the intervention plan data that thirty-six jurisdictions (just under half) were providing.
- ❖ David Cotton replied that they were using the data for process monitoring. The speaker noted that his state does not have a system, and so he and a colleague worked many hours to translate their data into reportable form for CDC. They were proud of their efforts, but were disappointed to learn that not all jurisdictions reported, rendering that data less-useful.

- ❖ Choi Wan said that the percentage of health departments who reported was from two weeks previously, and they had received information since then. He agreed that more health departments must report process monitoring data.
- ❖ David Cotton added that CDC is actively involved with the data that they got on intervention plans and that they plan to do the same thing with process monitoring. They recognize that there are going to be unique aspects to the data received and hope for fuller data the “next time around.” The numbers are too small at this point to do something that is interpretable, and process monitoring data are too new to have analyzed.
- ❖ Choi Wan said that CDC had given the jurisdictions feedback on their intervention plans. In the area of process monitoring, they are examining the discrepancies between the data provided and what the Guidance suggested. In the long run, they hope to help keep the information good and the turnaround time fairly fast.
- ❖ An inquiry was posed as to whether CDC was sure that the CDC health department management system would not one day be mandated for all jurisdictions. Since this participant was developing his own system, he hoped to avoid spending time and resources developing a system, only to find out later that, for standardization purposes, all states would be required to follow CDC’s system.
- ❖ Choi Wan stated that the CDC system would not replace any systems that have been developed or are under development by health departments. They see the need for a CDC-created system and the accompanying TA, so they want to provide the product to states that do not have the capacity to develop their own systems.
- ❖ Jaime Altamirano added that CDC has a problem with epi-profiles, which are different from one state to another, so it is not possible to compare profiles across states. Whatever system is used, he stressed that there must be a core of standard information that states have in common.
- ❖ It was noted that there are two pieces to doing work on the web. The web interface which Mari Gasiorowicz showed them is separate from the underlying database. CBO’s work with the interface, and the database is often determined by the health department’s IT department.
- ❖ An participant commented that in Florida, they were able to complete their intervention plan data. It took hours, because they do not have a system, and they were still collecting data using their old, pre-Evaluation Guidance method. This method did not yield the cross-tabs that CDC wants, but they worked to create them. The data is flawed, for

instance, a contract written to reach “Black men” in the intervention meant that the health department had to guess the ages, risks, and other attributes of that population for that study. She hoped that those data would not be used to analyze activities in Florida. Reporting the process monitoring data, then, seemed pointless because they had not collected data the right way and the data would have meant nothing. She hoped that CDC would acknowledge that states without a system would not be able to provide good data, and advocated waiting for a good system rather than filling out the forms for the sake of satisfying the requirement.

- ❖ David Cotton asked whether they now have a system in place for data collection that will, next year, yield data that will be better. She replied that they would for the next progress report because of the new cycle of providers. David Cotton asked whether other jurisdictions were in similar situations, having to wait for cycles to be complete before good process monitoring data can be collected.
- ❖ A participant noted that developing a database depends on variables, and she wondered if CDC staff could say when the variable definitions would stop changing.
- ❖ Jaime Altamirano sympathized with the question, adding that taxonomy and definitions can be a large problem. The simple definitions from the Guidance were questionable in certain jurisdictions, and standardization was difficult. CDC gets regular, new mandates from Congress about how to address certain populations, and they have to adapt to those changes.
- ❖ Hope Cassidy-Stewart noted that in Maryland, they collect more specific information that what CDC wants, so that when taxonomic changes occur, they can collapse their data differently. The interface that the CBO’s use does not have to be what is sent to CDC.
- ❖ A participant asked that the CDC data collection and reporting system give states flexibility. With flexibility in collecting process data and in, for instance, individual client-level data, states can work with their CBO’s to collect data in the best way possible and still be able to be aggregate into the appropriate categories.
- ❖ Another speaker advocated for flexibility in such areas as cross-tabbing referrals. The system should serve the jurisdiction and also CDC’s expanded requirements, with the potential for adding variables that are not part of the Guidance, but which states may need and use.

David Cotton then led the group in a brainstorming session of characteristics of an ideal data collection and management system. The participants generated the following list:

- ☐ The system should accommodate storing data.
- ☐ Data should be reportable at a local level.
- ☐ Quality assurance is an important component – what is entered at the micro level affects the macro level.
- ☐ GIS information is very useful to CBO's.
- ☐ Reports should be able to be sorted by client code, et cetera, in the interface. Also look at how client files are arrayed in the system.
- ☐ The system should coordinate with HRSA, SAMHSA, CSAP, and other agencies to which CBO's are required to report so that separate collecting of information does not have to occur.
- ☐ The system should have the ability to trace clients across agencies. This is difficult, but critical to see how the client's treatment goes and to trace the impact of various agencies on a given client. Data collection for outreach, in particular, is a difficult issue. One state has index cards that outreach workers use for notes, and outreach has the widest variability in data. There are ways to code client data that will assure confidentiality.
- ☐ Multiple and simultaneous users should have access to the system, and the system should be able to support them.
- ☐ Many CBOs do not have the capacity to use a web-based system, so CDC should develop a product that is compatible across machines with very basic technology levels.
- ☐ Perhaps ILI and group-level interventions should be the focus, and outreach can come later in the priority use of resources.